

Ongoing evolution of proposal reviews in the Spitzer warm mission

Lisa J. Storrie-Lombardi^{*a}, Suzanne R. Dodd^b, Nancy A. Silbermann^a, L. M. Rebull^a, Seppo Laine^a,
Megan Crane^a, John Stauffer^a, Lee Armus^a

^aSpitzer Science Center, California Institute of Technology, MC 314-6, 1200 E. California Blvd,
Pasadena, CA, USA 91125;

^bJet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA, USA 91109

ABSTRACT

The Spitzer Space Telescope is executing the seventh year of extended warm mission science. The cryogenic mission operated from 2003 to 2009. The observing proposal review process has evolved from large, week-long, in-person meetings during the cryogenic mission to the introduction of panel telecon reviews in the warm mission. Further compression of the schedule and budget for the proposal solicitation and selection process led to additional changes in 2014. Large proposals are still reviewed at an in-person meeting but smaller proposals are no longer discussed by a topical science panel. This hybrid process, involving an in-person committee for the larger proposals and strictly external reviewers for the smaller proposals, has been successfully implemented through two observing cycles. While people like the idea of not having to travel to a review it is still the consensus opinion, in our discussions with the community, that the in-person review panel discussions provide the most satisfying result. We continue to use in-person reviews for awarding greater than 90% of the observing time.

Keywords: Spitzer Space Telescope, proposal peer review, time allocation committee

1. INTRODUCTION

The Spitzer Space Telescope was launched in 2003 as NASA's Great Observatory for Infrared Astronomy. Extended mission operations began July 28, 2009 after successful completion of the five and a half year cryogenic mission. The Spitzer Science Center (SSC) has responsibility for the selection of observing proposals for the Spitzer mission. During the prime mission the proposal selection process worked in a similar fashion for all three of NASA's operating Great Observatories: Spitzer, Hubble, and Chandra. Broad science review panels organized by science topic reviewed the proposals. The individual reviewers graded the proposals in advance and then the panels met in person to discuss the proposals and provide a final ranking for the proposals within the panel, followed by a Time Allocation Committee (TAC) meeting that provides further review of the largest proposals. To handle 600 – 1200 proposals, the review meetings last a week and involve substantial investments in time from the science center (~35 people) and the community (100+ reviewers) as well as financial resources supporting travel of the reviewers and the meeting facilities. More details of the full Spitzer proposal selection process during the prime mission are discussed in Laine et al. (2006)¹.

Spitzer's extended operations phase, the "warm mission," currently operates with a staff of 25% of the prime mission level while continuing to execute more than 7000 hours of science observations per year. The warm mission operations model introduced support for substantially larger observing programs of greater than 500 hours called "Exploration Science" programs. With more time allocated to larger programs, the observatory can continue executing the same number of hours of science but in a substantially reduced number of individual programs that require support. During the prime mission the SSC supported ~ 250 programs per year and in the warm mission the SSC supports ~ 65 programs per year. One of the largest non-labor annual expenses during the prime mission for the SSC was the annual proposal review meeting, costing ~\$300,000. For the warm mission we implemented a review process where broad topical science panels still discussed the proposals, but it was done in a half-day telecon instead of an in-person meeting. The panel review telecons are followed by a face-to-face meeting of the TAC, with the TAC providing the final review for the largest proposals. This was the process for observing Cycles 6 – 10 and costs substantially less to support as only ~10 people need to travel to the TAC meeting.

^{*}lisa@ipac.caltech.edu; phone 1-626-395-8665; ssc.spitzer.caltech.edu

In 2014 the process for the solicitation and selection was further compressed due to changes in the mission funding profile. Specifically, we changed the review process for the smaller proposals. The large proposals are still reviewed by topical science panels that meet in person to discuss the proposals but no panel discusses the smaller proposals. Instead, individual external reviewers read and grade the smaller proposals and provide brief written comments. A ranked list is created by the SSC from the grades submitted by the individual reviewers. This new process has been utilized for Cycles 11 – 12 and will again be used in Cycle 13. We discuss here our experience with it and the pros and cons of the various strategies we have utilized for proposal reviews.

2. THE STATUS QUO: IN-PERSON REVIEW MEETINGS

Reviewing proposals with an in-person review meeting is an excellent way to provide a fair, reasonable process for selecting the best science to execute. Observatory staff put substantial effort into making sure the review process is as fair and unbiased as possible. It is critically important that the astronomical community has confidence in the review processes. Anyone who has participated in the review process understands that the goal is not to converge on a single unique answer, but instead the review process should result in any one of several “good” answers if done properly. For observatories that receive hundreds of proposals, and particularly those with international partnerships and active communities on multiple continents, the logistics of running an in-person meeting in a single location may be the simplest way to accomplish the review. A 12-hour time difference for participants is inconvenient for a one-hour meeting but completely impractical for managing a remote review with people spread in multiple time-zones across the planet.

On the downside, in-person meetings are expensive to run in terms of dollars, people resources and carbon footprint. Managing reviewer conflicts of interest necessitates creating review panels that are not always constituted optimally for every scientific topic or proposal. Traveling any distance to the review can add multiple days to the time commitment for the reviewers, in addition to the time spent preparing for the meeting and the meeting itself. A good success rate in recruiting reviewers for an in-person review is about fifty-percent (see section 6).

3. TELECON REVIEW MEETINGS

Prior to the start of the warm mission, a detailed review of the prime mission preliminary grades submitted by the panel reviewers, compared to the final proposal grades after the panel discussions, showed that the panel discussions at the review meeting had a consistent and material impact on the final ranked list of observing proposals (Storrie-Lombardi et al. 2008)². The typical effect of the panel discussion was a 25 – 30% change in the proposals that would have been selected. This led the SSC to not drop the discussion of the proposals as part of the review process at the start of the warm mission. The alternative that was considered was to use the initial grades submitted by the reviewers to make the final ranked list for some or all categories of proposals.

The review process by telecon works quite well because it follows a regular format that avoids most of the issues with meetings held with all participants on the phone. The general flow is:

1. Primary reviewer introduces the proposal and makes comments.
2. Secondary reviewer adds comments.
3. Comments requested from other panelists and general discussion.
4. Regrade the proposal and move on to the next one.

One of the concerns we had with not having face-to-face meetings was how it would impact reviewers participating in the process for the first time. We typically have a few post-docs as reviewers during each cycle and they have all been active participants in the process. We see no evidence that they do not participate fully even though most have not previously been part of a large proposal review.

The primary elements that impact how smoothly the review discussions by telecon work are the:

1. Number of proposals per panel
2. Breadth of science reviewed by the panel

These are important factors in any proposal review process but are particularly important for a review meeting limited to half a day. The Spitzer review panel telecons are scheduled for 3.5 hours (9:00am – 12:30pm Pacific Time) so that reviewers in the Americas can complete the review within normal work hours. We have also had a small number of reviewers participate from Europe and Hawaii.

Number of Proposals Reviewed: The number of proposals per panel is the most important factor in terms of the fidelity of the initial grades submitted by the reviewers. While most reviewers are conscientious about reading and grading proposals, it is extremely difficult for any individual to provide a ranked list of 50 – 100 proposals, which is now typical in many large observatory proposal reviews. For the warm mission we receive ~150 proposals/cycle. In the first warm mission cycle (Cycle 6) we assigned 30 – 35 proposals per panel; for Cycle 7 onwards it was limited to 20 – 25 proposals per panel. This meant more panels with fewer proposals each. Discussing 15 proposals in a half-day telecon, after triaging the lowest ranked proposals based on the preliminary grades, proved to be the optimal number to discuss.

Breadth of Science: Given the breadth of science that is proposed to Spitzer (see Figure 1) there will never be an unconflicted expert on the panel for every single proposal. It is also the case that experts in the proposed science can be the most critical reviewers and/or the most avid champions for particular science. Proposals written for a broad audience of reviewers, not just for experts, are always appreciated by the review panels. The breadth of science covered by the panel does affect the dynamics of the review, but there is no clear guideline we can use to determine what mix of proposals is too broad or too narrow. Avoiding conflicts or the appearance of conflicts is also a major factor in populating the review panels and assigning proposals to reviewers. Early in the warm mission we had fewer, broader panels. The move to more panels with fewer proposals per panel allows us to do a better job of matching experts to proposals while minimizing conflicts. Having parallel panels to handle similar science is critical so that we can have active proposers participating as reviewers.

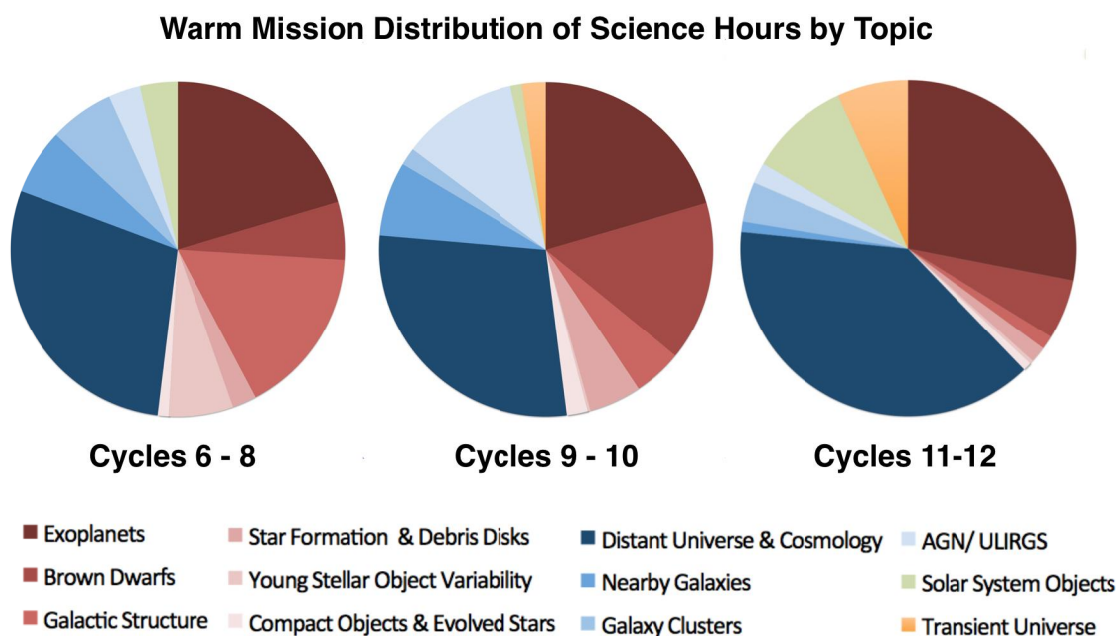


Figure 1. Spitzer supports a very broad and evolving range of scientific investigations. While studies of exoplanets and the distant universe command large amounts of time in each Cycle during the warm mission, other fields ebb and flow and new fields emerge, e.g., studies of the extragalactic transient universe starting in Cycle 10.

The primary downside to running more panels is that this puts a larger load on the SSC teams supporting the process. Supporting eight half-day telecons over two weeks, followed by a TAC meeting three weeks later, is an intense effort. The upside is that the reviewers have fewer proposals to read. As discussed in section 3 we kept the panel reviews for the warm mission due to the fact that there was a substantial change (25 – 30%) in what would have been selected for

Cycles 3 – 5 in the cryogenic mission if we had gone strictly with the preliminary grades with no panel discussion follow-up. We saw similar numbers in Cycle 6, when we moved to the panel telecons, but in Cycles 7 – 8 the magnitude of the change dropped to about 15%. This is summarized in Figure 2. While some panels have substantial changes in the final ranked list in every cycle, regardless of whether the discussion is done in person or via telecon, there is a noticeable drop in the mean level of the change in Cycle 7 – 8 compared to the previous four proposal reviews. We believe this is most likely due to limiting the number of proposals that each reviewer had to read to less than 25. Fewer proposals leads to demonstrably higher fidelity initial assessments by each reviewer.

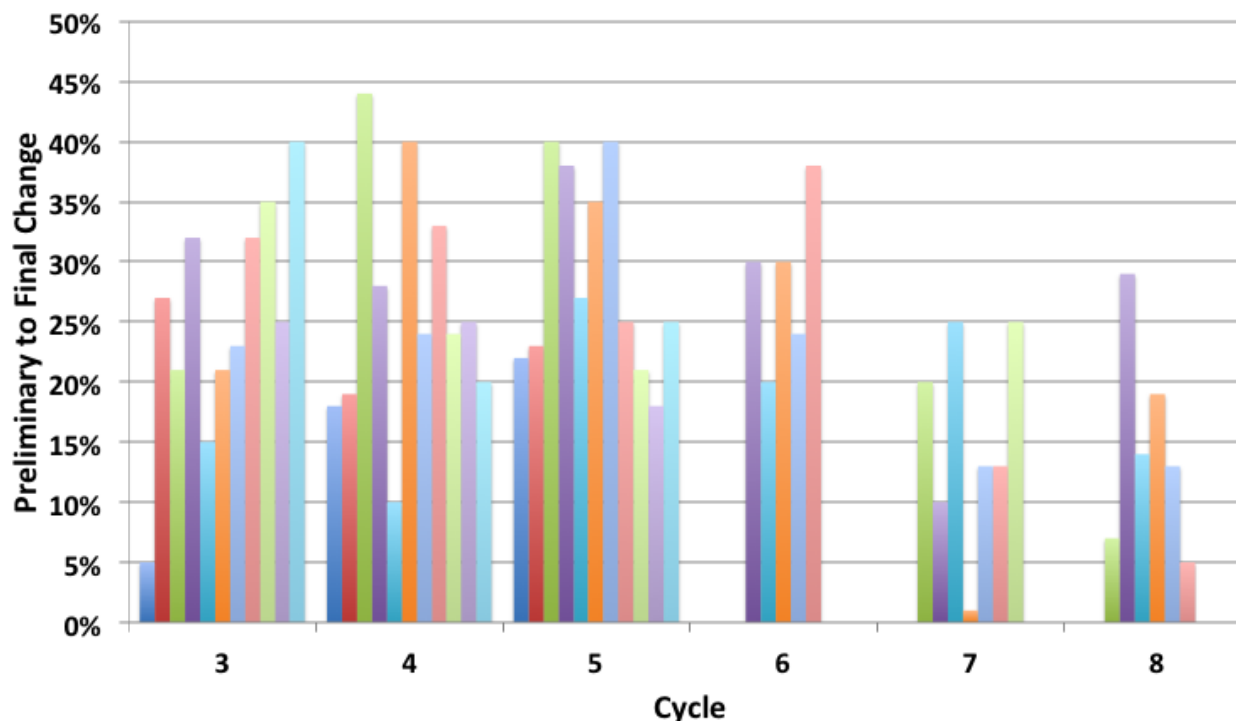


Figure 2. The percentage change in the proposals that were selected based on the final panel grades vs. the preliminary grades submitted prior to the panel discussions are shown for six observing cycles. Each bar represents an individual panel. There are no trends that correlate with the science discussed by the panel. Cycles 3 – 5 are the last three cycles of the cryogenic mission where we had 11 panels with 40 – 80 proposals/panel and all of the panels met in person. Cycles 6 – 8 are the first three cycles of the warm mission where we had 5 – 7 panels. Cycle 6 had 30 – 35 proposals/panel and Cycle 7 – 8 had no more than 25 proposals per panel. All but one of the panels in Cycles 6 – 8 were done via telecon. While some panels have substantial changes in the final ranked list in every cycle there is a noticeable drop in the mean level of the change in Cycle 7 – 8 compared to the previous four proposal reviews. We believe this is most likely due to limiting the number of proposals that each reviewer read to no more than 25.

4. EFFECTIVE PANEL TIME MANAGEMENT

It has been our experience that the mix of proposal sizes reviewed by a panel does not have a large impact on how smoothly the review panel works. Reviewers try to focus more attention on the larger proposals but will also carefully discuss the pros and cons of a one-hour proposal. For Cycle 10 we implemented a new strategy for reviewing the small proposals, i.e. those requesting less than 10 hours. The small proposals made up 30% of the proposals submitted by number but less than 4% of the total hours requested. The reviewers graded all of the proposals in advance but during the review telecon only discussed proposals requesting greater than 10 hours. The SSC used the preliminary grades submitted by the panelists to make an accept/reject list of the small proposals. The majority of the panels agreed with the accept/reject list though they were able to “rescue” and discuss any small proposals that were on the reject list. For Cycle 10 only two small proposals were rescued for discussion from the initial triage list. By not spending time discussing the small proposals the panels were able to discuss the larger proposals in greater depth.

For Cycle 10 we had no more than 25 proposals per panel. Almost all of the proposals requesting greater than 10 hours were discussed at the telecon. Only two of the larger proposals that were initially below the 50th percentile, based on the preliminary grades, received observing time. This suggests that rather than accepting/rejecting just the small proposals based on the preliminary grades we could triage half of all of the proposals, regardless of size, based on the preliminary grades. While only two proposals that had preliminary averages below the 50th percentile received time, they were proposals with large dispersions in the grades. So a strategy that aims to triage at the 50th percentile, but moves proposals with very high dispersions in the grades into the discussion list, is a fair plan for best utilizing limited resources.

For the proposals requesting greater than 100 hours, the panel discussion and later discussion by the TAC are critical in determining the final ranked list of recommended proposals. While the preliminary grades for the larger proposals are a good indicator of which proposals should be discussed, there is much more movement in grades for the larger proposals after the panel and/or TAC discussion. Larger programs often have larger teams and it is more difficult to get all competing proposals into the same panel, due to conflicts, so having the TAC review all of the highly ranked proposals from the panels continues to be the optimal method for the allocation of the majority of the observing time.

5. REINVENTING THE REVIEW PROCESS AGAIN

In 2014 the Spitzer mission was extended for two more years of operations but the extension was not confirmed until July and the nominal mission end date was September 30. The operations funding was also reduced. We needed to get the Cycle-11 observations selected as rapidly as possible to refill the scheduling pool. This led us to modify the review process yet again. We elected to have an in-person panel review, followed immediately by the TAC meeting, that only reviewed the larger proposals (> 100 hours). This is the same overall process we used in the cryogenic mission but with a much smaller total number of proposals, e.g., 60 instead of several hundred. Having the panel telecons with a 2 – 3 week break before the TAC, as we did in Cycle 10, would have added more than a month to the review timeline.

The Cycle-11 review was a 2.5 day meeting with five panels meeting the first day and forwarding onto the TAC the highest ranked proposals. The TAC discussed those proposals on the second day and on the third half-day determined the final recommended list of programs. In talking with Cycle-11 panelists who had served as reviewers on the panel telecons, while they appreciated not having to travel to the review when done by telecon, they still preferred the quality of the interactions with everyone in the same room. We did have a couple of people participate remotely, via Skype, so they had a video presence as part of the panel. Ninety percent of the Cycle-11 time was awarded to programs requesting more than 100 hours and all were reviewed at this meeting.

For the less than 100-hour proposals we used a completely independent set of external reviewers. They each received no more than 20 proposals to read, grade, rank, and provide comments for, and that completed their service. The advantage here is that proposals can be better targeted to expert reviewers because there are no issues with potential conflicts with other panel members, and that each proposal has more reviewers than they would with the traditional panels (thereby averaging over any outliers). The disadvantage is that disparate opinions from the reviewers are not resolved in a panel discussion. The SSC merges the grades and comments to create the final ranked list of proposals. We have consulted with individual reviewers in a few cases to clarify comments on proposals with a wide dispersion in the grades. Cycle 12 only solicited proposals less than 100 hours so the entire review was again done with external reviewers.

Given what we learned in Cycles 7 – 8, that reducing the number of proposals that reviewers have to read leads to a result that is less likely to change when discussed by a group, we believe this is a scientifically valid way to complete the review given our resources. Selecting less than 10% of the observing time using solely external reviewers while maintaining the more typical review meeting for the majority of the time is a good compromise. A substantial advantage to having no travel or panel telecon requirement is that people are much more likely to say “yes” to the request to serve as a reviewer, which we discuss in the next section.

6. RECRUITING REVIEWERS

Recruiting reviewers is perhaps the most demanding part of running a proposal review. The success rate in getting reviewers for the in-person and telecon reviews was typically 45 – 55%. Getting reviewers to serve via telecon is slightly, but not substantially, easier than getting them to commit to an in-person meeting. Far fewer people cancel at the last minute for the in-person reviews and in our experience they do so only because of some major personal or

professional issue. For the cryogenic mission the number of reviewers that could not show up was zero to one per review. For the telecon reviews it was closer to 5%.

For our current Cycle-13 review we have recruited an in-person TAC and it has proved the most difficult one to put together (31% success rate). We believe the reason for this is the timing of the review in late July – the peak season for summer science conferences and family vacations. As in Cycle 11, the dates of the review are driven primarily by the timing of the approval process for extending operations.

Recruiting external reviewers with no travel or telecon commitment has a much higher success rate. As discussed in section 5 these reviewers receive no more than 20 proposals to read, grade, rank, and for which they write brief comments. The success rate for recruiting these reviewers is 65 – 75% and with a much shorter lead-time before the review as we do not start recruiting the external reviewers until we have populated the panels and TAC for the in-person meeting.

7. CONCLUSIONS

Holding the panel review discussions by telecon, rather than with a face-to-face meeting, works extremely well if the total number of proposals in the review is less than 200. The process could be scaled to handle a larger number of proposals by running more panels simultaneously, but based on our experience it is unlikely this would work well for reviews handling ~1000 proposals. By limiting the number of proposals per panel to no more than 25, and running more panels, the initial grades submitted by the reviewers have high enough fidelity to be able to triage half of the proposals before the review panels or TAC meet. Allowing exceptions to the 50% triage for proposals with middling preliminary grades, but high dispersions, is also recommended.

The topical panel and TAC discussions, whether held via telecon or in a face-to-face meeting, are still critical for determining the final ranked list of the larger proposals. In particular, a face-to-face TAC meeting for the final selection is still the desired method based on discussions with our community. The final order of the larger proposals that start above the 50th percentile does change substantially based on the review discussion. While it would be ideal to be able to have panel discussions for all of the proposals, a process utilizing strictly external, independent reviewers for recommending the observing time awards for less than 10% of the observing time provides a good solution in a very resource-limited operations environment.

REFERENCES

- [1] Laine, S., Silbermann, N.A., Rebull, L.M. and Storrie-Lombardi, L.J., "Spitzer Space Telescope proposal process," Proc. SPIE 6270, 29 (2006).
- [2] Storrie-Lombardi, L.J., Silbermann, N.A., Rebull, L.M., Laine, S., and Crane, M., "Review Rankings: The influence of reviewer discussions on proposal selection," Proc. SPIE 7016, 221 (2008).